

RELAXED CONCENTRATED MLE FOR ROBUST CALIBRATION OF RADIO INTERFEROMETERS

V. Ollier^{*†}, M. N. El Korso[†], R. Boyer[‡], P. Larzabal^{*} and M. Pesavento[§]

^{*} SATIE, UMR 8029, ENS Cachan, Université Paris-Saclay, Cachan, France

[†] LEME, EA 4416, Université Paris-Ouest, Ville d'Avray, France

[‡] L2S, UMR 8506, Université Paris-Sud, Gif-sur-Yvette, France

[§] Communication Systems Group, Technische Universität Darmstadt, Darmstadt, Germany

ABSTRACT

In this paper, we investigate the calibration of radio interferometers in which Jones matrices are considered to model the interaction between the incident electromagnetic field and the antennas of each station. Specifically, perturbation effects are introduced along the signal path, leading to the conversion of the plane wave into an electric voltage by the receptor. In order to design a robust estimator, the noise is assumed to follow a spherically invariant random process (SIRP). The derived algorithm is based on an iterative relaxed concentrated maximum likelihood estimator (MLE), for which closed-form expressions are obtained for most of the unknown parameters.

Index Terms— Calibration, Jones matrices, robustness, SIRP, relaxed concentrated maximum likelihood

1. INTRODUCTION

The new generation of radio telescopes, such as the low frequency array (LOFAR) [1] and the square kilometre array (SKA) [2], are characterized by a large number of receiving elements, large collecting area and dynamic range, wide field of view, high sensitivity and resolution, huge amount of measurement data, etc., which entails a certain number of scientific challenges. In radio interferometry [3], one of the most important challenges is notably the calibration [4].

Calibration involves the estimation and the correction of different unknown perturbations introduced along the signal path, e.g., due to the environment (atmosphere, ionosphere) or the artifacts in the instruments (electronic gain, bandpass, station beam shape and orientation, sidelobe contamination, etc.). All these physical corruption effects, which may be direction dependent [5], are involved in the radio interferometer measurement equation and can be modeled with the help of Jones matrices [6, 7]. Besides, many faint sources are present in radio interferometric data and can be considered as outliers

in the calibration procedure, leading to deviations from the commonly assumed Gaussian noise model [8, 9].

To overcome these drawbacks, we intend to robustify the calibration scheme by using a wider distribution class than the Gaussian one, to model the noise. In doing so, we do not specify precisely the noise distribution, unlike [10] where the Student's t-distribution is considered, and we use a broad class of distributions gathered under the so-called spherically invariant random distribution [11, 12]. A spherically invariant random process (SIRP) is described as the product of a texture parameter: a positive random variable, and a speckle component: a Gaussian process, resulting in a two-scale compound-Gaussian distribution. Under SIRP noise, the maximum likelihood (ML) method can be used to estimate the unknown parameters [13]. In our case, to obtain closed-form expressions and to reduce the computational complexity of our problem, the ML estimates are derived in an iterative way with a sequential updating procedure [14]. This is the iterative concentrated ML technique. However, a numerical optimization process still needs to be performed and can be computed efficiently for instance with the Levenberg-Marquardt (LM) solver [15, 16].

In this paper, we use the following notation: symbols $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ denote, respectively, the transpose, the complex conjugate and the Hermitian operator. The Kronecker product is represented by \otimes , $E\{\cdot\}$ denotes the expectation operator and $\text{bdiag}\{\cdot\}$ is the block-diagonal operator. The trace and determinant operators are, respectively, referred to by $\text{tr}\{\cdot\}$ and $|\cdot|$. Finally, the symbol \mathbf{I}_B represents the $B \times B$ identity matrix and $\text{vec}(\cdot)$ stacks the columns of a matrix on top of one another.

2. DATA MODEL

Let us consider D signal sources impinging on a station of M antennas. Each electromagnetic plane wave is observed by M antennas and can be decomposed as two orthogonal polarization directions (x, y) . Each antenna is composed of two receptors, which are sensitive to a particular polarization

This work was supported by the following projects: MAGELLAN (ANR-14-CE23-0004-01) and by the iCODE institute, research project of the IDEX Paris-Saclay.

[6]. Thus, the measured voltage at the p -th antenna due to the i -th signal source impinging on it is written as [6, 7]

$$\bar{\mathbf{v}}_{i_p} = \mathbf{J}_{i_p}(\boldsymbol{\theta}) \mathbf{s}_i \quad (1)$$

where the relation between each incoming radiation $\mathbf{s}_i = [s_{i_x}, s_{i_y}]^T$ and the generated voltage at each antenna $\bar{\mathbf{v}}_{i_p} = [v_{i_{p_x}}, v_{i_{p_y}}]^T$ is given by a 2×2 Jones matrix $\mathbf{J}_{i_p}(\boldsymbol{\theta})$, parametrized by the unknown vector $\boldsymbol{\theta}$. The Jones matrix accounts for the different perturbations introduced on the path from the i -th source to the p -th sensor. Hence, for a given source-antenna pair, we measure two output signals, i.e., one for each polarization. Since each Jones matrix is associated with a source-antenna pair, the total number of Jones matrices is DM .

Typically, in radio astronomy, signals collected by a pair of antennas (p, q) , i.e., two pairs of output signals, are correlated. Then, different crosscorrelation measurements, called visibilities, are computed for different antenna pairs, with a specific baseline. The total number of antenna pairs is $B = \frac{M(M-1)}{2}$ and for a given (p, q) antenna pair, the 2×2 visibility matrix, in the noiseless case, is denoted by $\tilde{\mathbf{V}}_{pq} = \mathbb{E}\{\bar{\mathbf{v}}_p \bar{\mathbf{v}}_q^H\}$ and written as

$$\tilde{\mathbf{V}}_{pq} = \sum_{i=1}^D \mathbf{J}_{i_p}(\boldsymbol{\theta}) \mathbf{C}_i \mathbf{J}_{i_q}^H(\boldsymbol{\theta}) \quad \text{for } p < q, \quad p, q \in [1, \dots, M], \quad (2)$$

in which, for the i -th source, $\mathbf{C}_i = \mathbb{E}\{\mathbf{s}_i \mathbf{s}_i^H\}$ is the 2×2 intrinsic source coherency matrix, known from prior knowledge. As one can notice, this equation is composed of the contributions from D discrete polarized calibrator sources ($D > 1$ to avoid calibration ambiguities [5]) with uncorelated radiations and the corresponding path effect modeled by the Jones matrices.

Using [17, p. 424], the vectorized form of (2) can be written as

$$\tilde{\mathbf{v}}_{pq} = \text{vec}(\tilde{\mathbf{V}}_{pq}) = \sum_{i=1}^D \mathbf{u}_{i_{pq}}(\boldsymbol{\theta}) \quad (3)$$

where $\mathbf{u}_{i_{pq}}(\boldsymbol{\theta}) = (\mathbf{J}_{i_q}^*(\boldsymbol{\theta}) \otimes \mathbf{J}_{i_p}(\boldsymbol{\theta})) \mathbf{c}_i$, in which $\mathbf{c}_i = \text{vec}(\mathbf{C}_i)$. To consider a more realistic scenario, we introduce a noise vector \mathbf{n}_{pq} for each antenna pair (p, q) such that the visibility vector becomes

$$\mathbf{v}_{pq} = \tilde{\mathbf{v}}_{pq} + \mathbf{n}_{pq}. \quad (4)$$

The full visibility vector \mathbf{x} of length $4B$ is given by

$$\mathbf{x} = \begin{bmatrix} \mathbf{v}_{12} \\ \mathbf{v}_{13} \\ \vdots \\ \mathbf{v}_{(M-1)M} \end{bmatrix} = \sum_{i=1}^D \mathbf{u}_i(\boldsymbol{\theta}) + \mathbf{n} \quad (5)$$

where $\mathbf{u}_i(\boldsymbol{\theta}) = [\mathbf{u}_{i_{12}}^T(\boldsymbol{\theta}), \mathbf{u}_{i_{13}}^T(\boldsymbol{\theta}), \dots, \mathbf{u}_{i_{(M-1)M}}^T(\boldsymbol{\theta})]^T$ and $\mathbf{n} = [\mathbf{n}_{12}^T, \mathbf{n}_{13}^T, \dots, \mathbf{n}_{(M-1)M}^T]^T$.

3. MAXIMUM LIKELIHOOD ESTIMATION

The D signal sources correspond to the brightest sources, while the weak ones are considered as noise. Therefore, outliers may appear and the Gaussian noise assumption may not be fulfilled [18], e.g., a Student's t -distribution may be used [10]. To cope with different noise distributions, specifically non-Gaussian noise modeling, and to achieve robust calibration w.r.t. outliers, we consider a SIRP, which is defined for each antenna pair as

$$\mathbf{n}_{pq} = \sqrt{\tau_{pq}} \mathbf{g}_{pq} \quad (6)$$

where the random variable τ_{pq} is positive and real. This power factor varies independently according to the antenna pair considered and, in the radar context, is called texture. The speckle component \mathbf{g}_{pq} is a complex zero-mean Gaussian process with an unknown covariance matrix $\boldsymbol{\Omega}$, i.e.,

$$\mathbf{g}_{pq} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Omega}) \quad \text{such that} \quad \text{tr}\{\boldsymbol{\Omega}\} = 1, \quad (7)$$

where the 4×4 covariance matrix $\boldsymbol{\Omega}$ is the same for all antenna pairs and a constraint is required on its trace to remove scaling ambiguities in model (6). Taking into account such noise model and assuming spatial independence between antenna pairs, the likelihood function is given by

$$f(\mathbf{v}_{12}, \dots, \mathbf{v}_{(M-1)M} | \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}) = \prod_{pq} \frac{1}{|\pi \tau_{pq} \boldsymbol{\Omega}|} \exp \left\{ -\frac{1}{\tau_{pq}} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}) \right\}, \quad (8)$$

with $\boldsymbol{\tau} = [\tau_{12}, \tau_{13}, \dots, \tau_{(M-1)M}]^T$ and $\mathbf{a}_{pq} = \mathbf{v}_{pq} - \tilde{\mathbf{v}}_{pq}$. In the r.h.s. of (8), the product is performed for each antenna pair so there are B elements in the product. The log-likelihood function is written as follows

$$\begin{aligned} \log f(\mathbf{v}_{12}, \dots, \mathbf{v}_{(M-1)M} | \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}) &= -4B \log \pi \\ &- 4 \sum_{pq} \log \tau_{pq} - B \log |\boldsymbol{\Omega}| - \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}). \end{aligned} \quad (9)$$

The proposed robust calibration scheme is based on an iterative ML algorithm [14, 19]. The principle is to optimize the log-likelihood function w.r.t. each unknown parameter, while fixing the others, leading to the so-called concentrated ML estimator. Furthermore, we use in the following a relaxed ML estimator for which the texture parameters are assumed unknown and deterministic. This choice is motivated by the fact that we aim to design a broad robust estimator w.r.t. the presence of outliers but also in order to avoid a misspecification of the probability density function of $\boldsymbol{\tau}$ (i.e., we do not need to specify the texture distribution, ensuring more flexibility). A closed-form expression can be obtained for each texture realization τ_{pq} and the speckle covariance matrix $\boldsymbol{\Omega}$. Generally, no closed-form expression can be obtained for $\boldsymbol{\theta}$, unless

assuming a specific linear modeling of this vector w.r.t. the noiseless visibilities.

1) Derivation of $\hat{\tau}_{pq}$: We take the derivative of the log-likelihood function in (9) w.r.t. τ_{pq} and equate it to 0, leading to

$$-\frac{4}{\tau_{pq}} + \frac{1}{\tau_{pq}^2} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}) = 0. \quad (10)$$

We then obtain the expression of the texture estimate,

$$\hat{\tau}_{pq} = \frac{1}{4} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}). \quad (11)$$

2) Derivation of $\hat{\boldsymbol{\Omega}}$: We take the derivative of the log-likelihood function w.r.t. the element $[\boldsymbol{\Omega}]_{k,l}$ of the speckle covariance matrix and equate it to 0. Using [20, p. 2741], we obtain

$$-B \text{tr} \{ \boldsymbol{\Omega}^{-1} \mathbf{e}_k \mathbf{e}_k^T \} + \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{e}_k \mathbf{e}_l^T \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}) = 0 \quad (12)$$

where the vector \mathbf{e}_k contains zeros except at the k -th position which is equal to unity. Using the permutation property of the trace operator, we obtain

$$-B \mathbf{e}_l^T \boldsymbol{\Omega}^{-1} \mathbf{e}_k + \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{e}_l^T \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}) \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{e}_k = 0. \quad (13)$$

Consequently,

$$\hat{\boldsymbol{\Omega}} = \frac{1}{B} \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{a}_{pq}(\boldsymbol{\theta}) \mathbf{a}_{pq}^H(\boldsymbol{\theta}). \quad (14)$$

Since we adopt here an iterative procedure with a concentrated ML scheme, we plug (11) into (14) leading to

$$\hat{\boldsymbol{\Omega}}^{j+1} = \frac{4}{B} \sum_{pq} \frac{\mathbf{a}_{pq}(\boldsymbol{\theta}) \mathbf{a}_{pq}^H(\boldsymbol{\theta})}{\mathbf{a}_{pq}^H(\boldsymbol{\theta}) (\hat{\boldsymbol{\Omega}}^j)^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta})} \quad (15)$$

where j represents the j -th iteration. To ensure uniquely identifiable noise parameters, as it was previously mentioned in (7), the estimate of $\boldsymbol{\Omega}$ needs to be, e.g., normalized by its trace, as

$$\hat{\boldsymbol{\Omega}}^{j+1} = \frac{\hat{\boldsymbol{\Omega}}^{j+1}}{\text{tr} \{ \hat{\boldsymbol{\Omega}}^{j+1} \}}. \quad (16)$$

3) Estimation of $\hat{\boldsymbol{\theta}}$: Estimating $\hat{\boldsymbol{\theta}}$ for a given $\boldsymbol{\Omega}$ and τ leads to

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left\{ \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{a}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{a}_{pq}(\boldsymbol{\theta}) \right\}. \quad (17)$$

Depending on the structure of the Jones matrices [21], a different procedure can be adopted to estimate $\boldsymbol{\theta}$. A particular parametrization is the non-structured case, where $\boldsymbol{\theta}$ is composed of the entries of all Jones matrices, which is considered in the following.

4. ESTIMATION OF $\hat{\boldsymbol{\theta}}$ FOR NON-STRUCTURED JONES MATRICES

The optimization in (17) may be computationally heavy and very slow in convergence. To overcome this drawback, we apply the expectation-maximization (EM) algorithm, as in [8] and [9]. Since we adopt the non-structured Jones matrices case (i.e., $\boldsymbol{\theta}$ is a collection of Jones matrices' elements), the vector $\boldsymbol{\theta}$ can be partitioned as

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_D^T]^T = [\boldsymbol{\theta}_{11}^T, \dots, \boldsymbol{\theta}_{1M}^T, \dots, \boldsymbol{\theta}_{D1}^T, \dots, \boldsymbol{\theta}_{DM}^T]^T \quad (18)$$

meaning that for the i -th source and the p -th antenna, we have $\mathbf{J}_{i_p}(\boldsymbol{\theta}) = \mathbf{J}_{i_p}(\boldsymbol{\theta}_{i_p})$ in which $\boldsymbol{\theta}_{i_p}$ denotes the parametrization of the path i - p .

The EM algorithm [22, 23] is an iterative procedure to approximate the ML estimation technique and reduce its computational cost. First, the E-step computes the conditional expectation of the complete data given the observed data and the current fit for parameters. Second, the M-step maximizes the log-likelihood function of the conditional distribution, previously computed. This may not result in a closed-form expression and requires a numerical optimization procedure. The Levenberg-Marquardt (LM) algorithm is notably particularly appropriate for non-linear problems. The E- and the M-steps are repeated until convergence or until the maximum number of iterations is reached. The complexity is reduced since the unknown parameter vector is partitioned over the sources and optimization is carried out w.r.t. to $\boldsymbol{\theta}_i$ instead of $\boldsymbol{\theta}$. This leads to single source sub-optimization problems of smaller sizes.

1) E-step: We introduce the complete data vector $\mathbf{w} = [\mathbf{w}_1^T, \dots, \mathbf{w}_D^T]^T$ where, for the i -th source, the $4B \times 1$ vector \mathbf{w}_i is given by

$$\mathbf{w}_i = \mathbf{u}_i(\boldsymbol{\theta}_i) + \mathbf{n}_i \quad (19)$$

such that $\mathbf{x} = \sum_{i=1}^D \mathbf{w}_i$. The noise vectors \mathbf{n}_i are supposed to be statistically independent such that $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \beta_i \boldsymbol{\Psi})$ where $\sum_{i=1}^D \beta_i = 1$ and $\boldsymbol{\Psi}$ is the covariance matrix of \mathbf{n} . The covariance matrix of each noise vector \mathbf{n}_{pq} is given by $\tau_{pq} \boldsymbol{\Omega}$. Making use of the independence property, we obtain $\boldsymbol{\Psi} = \text{bdiag}\{\tau_{12} \boldsymbol{\Omega}, \dots, \tau_{(M-1)M} \boldsymbol{\Omega}\}$ and the covariance matrix of \mathbf{w} is given by $\boldsymbol{\Xi} = \text{bdiag}\{\beta_1 \boldsymbol{\Psi}, \dots, \beta_D \boldsymbol{\Psi}\}$.

Using [24, p. 36], we obtain the conditional expectation of the complete data $\hat{\mathbf{w}} = \mathbb{E}\{\mathbf{w}|\mathbf{x}; \boldsymbol{\theta}, \tau, \boldsymbol{\Omega}\}$, in the jointly Gaussian case, that is,

$$\hat{\mathbf{w}}_i = \mathbf{u}_i(\boldsymbol{\theta}_i) + \beta_i \left(\mathbf{x} - \sum_{l=1}^D \mathbf{u}_l(\boldsymbol{\theta}_l) \right). \quad (20)$$

2) M-step: Once $\hat{\mathbf{w}}$ is evaluated, $\boldsymbol{\theta}_i$ is estimated through

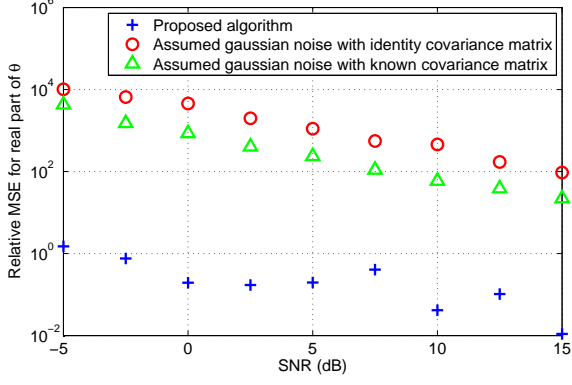


Fig. 1. Evolution of the relative MSE of a given unknown parameter as a function of the SNR.

optimization. Independence of \mathbf{w}_i leads to

$$f(\hat{\mathbf{w}}|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}) = \prod_{i=1}^D \frac{1}{|\pi\beta_i\Psi|} \exp \left\{ -\left(\hat{\mathbf{w}}_i - \mathbf{u}_i(\boldsymbol{\theta}_i)\right)^H (\beta_i\Psi)^{-1} \left(\hat{\mathbf{w}}_i - \mathbf{u}_i(\boldsymbol{\theta}_i)\right) \right\}. \quad (21)$$

For the i -th source, the cost function to minimize is given by $\zeta_i(\boldsymbol{\theta}_i) = \left(\hat{\mathbf{w}}_i - \mathbf{u}_i(\boldsymbol{\theta}_i)\right)^H (\beta_i\Psi)^{-1} \left(\hat{\mathbf{w}}_i - \mathbf{u}_i(\boldsymbol{\theta}_i)\right)$. At the $(h+1)$ -th iteration of the LM-like algorithm, we have:

$$\boldsymbol{\theta}_i^{h+1} = \boldsymbol{\theta}_i^h - (\nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_i}^T \zeta_i(\boldsymbol{\theta}_i) + \lambda \mathbf{I}_{4M})^{-1} \nabla_{\boldsymbol{\theta}_i} \zeta_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^h}. \quad (22)$$

Proposed algorithm:

input : $D, M, B, \mathbf{C}_i, \beta_i, \mathbf{x}$

output : estimate of $\boldsymbol{\theta}$

initialize: $\hat{\boldsymbol{\Omega}} \leftarrow \boldsymbol{\Omega}_{\text{init}}, \hat{\boldsymbol{\tau}} \leftarrow \boldsymbol{\tau}_{\text{init}}$

while stop criterion *unreached* **do**

- 1 $\hat{\boldsymbol{\theta}}_i \leftarrow \boldsymbol{\theta}_{i,\text{init}}, i = 1, \dots, D$
 - while** stop criterion *unreached* **do**
 - 2 E-step: $\hat{\mathbf{w}}_i$ obtained with (20)
 - 3 M-step: $\hat{\boldsymbol{\theta}}_i$ obtained with (22)
 - end**
 - 4 Obtain $\hat{\boldsymbol{\Omega}}$ with (15) then (16)
 - 5 Obtain $\hat{\boldsymbol{\tau}}$ with (11)
 - end**
-

5. NUMERICAL SIMULATIONS

In the following simulations, we consider $D = 2$ sources and $M = 16$ antennas. Therefore, the number of antenna pairs is $B = 120$, each providing a 4×1 observation vector, and there are 128 unknown parameters to estimate, corresponding to

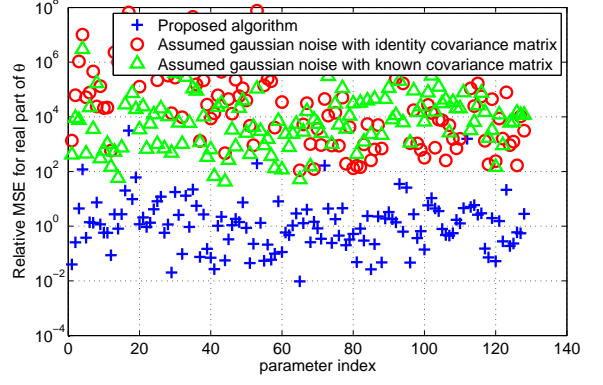


Fig. 2. Relative MSE of the 128 unknown parameters for a given SNR.

the entries of the Jones matrices. The source coherency \mathbf{C}_i is defined thanks to the Stokes parameters [6, 7], which represent the polarization state of the i -th signal source considered. The number of Monte Carlo runs is set to 100.

In Fig. 1, we plot the relative mean square error (MSE) vs. (residual) signal-to-noise ratio (SNR), for a given parameter representative of the overall behavior (by residual SNR, we mean the SNR computed using only the off-diagonal terms of the covariance matrix since the diagonal terms have been deleted, i.e., equivalently $\tilde{\mathbf{V}}_{pp}$ is not considered in the observation model (2)). We represent the following cases: i) the proposed algorithm as exposed in the pattern (blue curve) which intends to propose a robust estimator, ii) the case when gaussian noise is assumed with a known covariance matrix (equivalently $\boldsymbol{\Omega}$ is known and $\boldsymbol{\tau}$ is set to a vector filled with ones during the whole estimation procedure (green curve)) and iii) the case when gaussian noise is assumed with an identity covariance matrix (equivalently $\boldsymbol{\Omega}$ is set to the identity matrix and $\boldsymbol{\tau}$ is filled with ones (red curve)). The two last cases correspond to Gaussian modeling, with spatially correlated and i.i.d. noise. For all cases, the observations are generated using the true noise covariance matrix, structured as described in (6). In Fig. 2, we plot the relative MSE of each unknown parameter for a SNR of 15 dB.

The lowest MSE is achieved with the proposed algorithm, which estimates iteratively both texture and speckle components and the Jones matrices for calibration. Such performance is due to the SIRP noise assumption which includes many various distributions. The calculations were performed without precising the distribution considered, thus ensuring robust calibration to the presence of outliers in the noise (faint sources). If we do not characterize precisely the noise (i.e., we do not take into account the probability density function of $\boldsymbol{\tau}$), we have less information (relaxed ML estimator) but we reach robustness. The two other curves do not take into account the noise model as presented in (6), leading to mis-specifications and poor accuracy in calibration.

6. CONCLUSION

In this paper, we propose a robust calibration algorithm of radio interferometers, where non-structured Jones matrices are used to model the different perturbations introduced along the signal path. The derived scheme is based on an iterative relaxed concentrated ML method, in which a SIRP noise model is introduced, without fixing the texture distribution. Closed-form expressions are obtained for the noise parameters (texture components and speckle covariance matrix) while the estimation of the unknown vector θ is performed via an optimization process. The computational complexity of the problem is reduced with the use of the EM algorithm and the partition per source.

REFERENCES

- [1] M. P. van Haarlem et al., “LOFAR: The LOW-Frequency ARray,” *Astronomy & Astrophysics*, vol. 556, no. A2, 2013.
- [2] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, “The Square Kilometre Array,” *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.
- [3] A. R. Thompson, J. M. Moran, and G. W. Swenson Jr., *Interferometry and synthesis in radio astronomy*, 2nd ed., John Wiley & Sons, 2001.
- [4] S. J. Wijnholds, S. van der Tol, R. Nijboer, and A.-J. van der Veen, “Calibration challenges for future radio telescopes,” *IEEE Signal Processing Magazine*, vol. 27, no. 1, pp. 30–42, 2010.
- [5] O. M. Smirnov, “Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects,” *Astronomy & Astrophysics*, vol. 527, no. A107, 2011.
- [6] J. P. Hamaker, J. D. Bregman, and R. J. Sault, “Understanding radio polarimetry. I. Mathematical foundations,” *Astronomy & Astrophysics Supplement Series*, vol. 117, no. 1, pp. 137–147, 1996.
- [7] O. M. Smirnov, “Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism,” *Astronomy & Astrophysics*, vol. 527, no. A106, 2011.
- [8] S. Yatawatta, S. Zaroubi, G. de Bruyn, L. Koopmans, and J. Noordam, “Radio interferometric calibration using the SAGE algorithm,” in *13th IEEE DSP Workshop*, Marco Island, FL, 2009, pp. 150–155.
- [9] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Lampropoulos, A. G. de Bruyn, L. V. E. Koopmans, and J. Noordam, “Radio interferometric calibration using the SAGE algorithm,” *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, 2011.
- [10] S. Kazemi and S. Yatawatta, “Robust radio interferometric calibration using the t-distribution,” *Monthly Notices of the Royal Astronomical Society*, vol. 435, no. 1, pp. 597–605, 2013.
- [11] E. Jay, *Détection en Environnement non-Gaussien*, Ph.D. thesis, Université de Cergy-Pontoise, 2002.
- [12] K. Yao, “Spherically invariant random processes: Theory and applications,” in *Communications, Information and Network Security*, V. K. Bhargava et al., Eds., pp. 315–331. Springer, 2003.
- [13] J. Wang, A. Dogandzic, and A. Nehorai, “Maximum likelihood estimation of compound-Gaussian clutter and target parameters,” *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3884–3898, 2006.
- [14] X. Zhang, M. N. El Korso, and M. Pesavento, “Maximum likelihood and maximum a posteriori direction-of-arrival estimation in the presence of SIRP noise,” in *Proc. of IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Shanghai, China, 2016.
- [15] K. Madsen, H. B. Nielsen, and O. Tingleff, “Methods for non-linear least squares problems,” Tech. Rep., Informatics and Mathematical Modelling, Technical University of Denmark, 2004.
- [16] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed., Springer Science & Business Media, 2006.
- [17] A.-J. van der Veen and S. J. Wijnholds, “Signal processing tools for radio astronomy,” in *Handbook of Signal Processing Systems*, pp. 421–463. Springer, 2013.
- [18] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, “Robust statistical modeling using the t distribution,” *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.
- [19] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, “Joint ML calibration and DOA estimation with separated arrays,” in *Proc. of IEEE Int. Conf. on Acoust., Speech, Signal Processing*, Shanghai, China, 2016.
- [20] A. Hjørungnes and D. Gesbert, “Complex-valued matrix differentiation: Techniques and key results,” *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, 2007.
- [21] J. E. Noordam, “The measurement equation of a generic radio telescope, AIPS++ implementation note nr 185,” Tech. Rep., 1996.
- [22] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [23] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions*, 2nd ed., John Wiley & Sons, 2008.
- [24] T. W. Anderson, *An introduction to multivariate statistical analysis*, 2nd ed., John Wiley & Sons, 1984.